A Bayes Monte Carlo Approach to Reducing Uncertainty in Source Characterization and Building Description for a Transient Pollutant Release in a Residential House

Pamela Reynolds
Energy Research Undergraduate Laboratory Research Fellowship Program
Cabrillo College
Lawrence Berkeley Laboratory
Berkeley, CA 94720

Aug 7, 2000

Prepared in partial fulfillment of the requirements of the Department of Energy, Energy Research Undergraduate Laboratory Research Fellowship under the direction of Dr. Michael D. Sohn and Dr. Ashok J. Gadgil in the Airflow and Pollutant Transport Group, Indoor Environment Department, Energy and Environmental Technologies Division at Lawrence Berkeley National Laboratory.

Participant: ____

Research Advisor: ____

Research Advisor: ____

Table of Contents

Abstract

A Bayes Monte Carlo approach is used to reduce uncertainty in source characterization and building description for a transient pollutant release in a residential house. Uncertainty distributions for input parameters were subjectively assigned and updated using synthetically generated airborne pollutant concentration data. Two hypothetical data collection scenarios were considered: data collected in all five of the rooms every five minutes, and data collected in one different room every five minutes. Uncertainty estimations are updated sequentially at each five minute interval. Input uncertainty in source characterization is significantly reduced in both scenarios. Reduction in uncertainty is considerable when data is collected in more than one room simultaneously. Location and characterization of the pollutant release illustrates a typical application of this method.

Presenter's Name: Pamela Reynolds
Mailing Address: 206 Sylvan Way
City/State/Zip: Felton, CA 95018
Phone: 831-335-8316
E-mail Address: pamitha@usa.net
School Author Attends: Cabrillo Community College

DOE National Laboratory Attended: Lawrence Berkeley National Laboratory
Mentor's Name: Dr. Michael D. Sohn
Phone: 510-495-2448
E-mail Address: mdsohn@lbl.gov

Mentor's Name: Dr. Ashok J. Gadgil
Phone: 510-495-4651
E-mail Address: ajgadgil@lbl.gov

Introduction

Indoor air quality has received increasing attention in recent years. In response to rising concern over energy usage, buildings are becoming more airtight, with ventilation accomplished through mechanical HVAC systems (Sohn and Small 1999). A number of illnesses have been traced to problems of indoor air pollution, including the use of chemicals without proper ventilation, contaminant spills, and sick building syndrome. Many airborne contaminants may be colorless gasses, invisible to the naked eye. As we are beginning to understand the complexities of airflow inside buildings, we have learned that the transport of an airborne chemical released indoors may be non-, or even counter-intuitive. Scientists have attempted to understand indoor airflow and pollutant transport through computer modeling, using inputs of measurements for the building description and source characterization and producing predictions for pollutant concentration in various rooms.

Traditional methods of modeling pollutant releases rely on data gathered that characterize the source and building before modeling begins. It is, however, sometimes necessary to be able to understand a pollutant release without having high quality data describing the source and the building. We may have only remote data, occupant accounts, or pollutant concentration data. The Bayes Monte Carlo (BMC) approach described here is a method of reducing uncertainty in source characterization and building description in a small building using existing data, often of a limited quantity and poor quality.

The Bayes Monte Carlo method has been demonstrated to significantly reduce uncertainty in source characterization and transport parameters in other environmental media (see, eg, Sohn et al. 2000). The BMC method is different from most other methods of pollutant transport model calibration. Traditionally, the model is first completed using all known data. If new measurements are obtained, the calibration must be recalculated. This can be computationally intensive if measurements continue to be made as the model is being refined. Additionally, considerable field measurements may be required before modeling work is begun.

The BMC approach is unique in that it allows for the systematic use of expert judgment when limited data is available. As new data is obtained, the influence of expert judgment diminishes. Because Bayesian data analysis is de-coupled from the model simulation, continued updating alongside data gathering does not require running numerous additional simulations. Best estimates of source characterization and building description can be continually generated as data comes in. These features make it well suited for our purposes in this study.

The objective of this paper is to demonstrate that a BMC updating approach is useful and valid for reducing uncertainty in source characterization and building description for an indoor pollutant release in a small residential building. The study will also examine to what extent the quantity and quality of field data improves our updated estimates. An application of this work is to be able to locate and characterize the source in the indoor release of a pollutant, using scant or unreliable field data. This information could be used to direct emergency response personnel in the event of an accidental release.

Approach

Description of the Bayes Monte Carlo Method

The Bayes Monte Carlo method consists of two parts. In the Prior Analysis uncertainty distributions are assigned to the building description and source characterization input parameters. At this stage of the analysis, the model user allows the incorporation of subjective judgment when limited input information is available. Many realizations of these distributions are sampled using a standard Monte Carlo procedure.

In the Posterior Analysis field data are compared to the set of model predictions to assess how well each prediction fits the data, estimated using a likelihood function. We assumed a gaussian likelihood function:

$$L(O_s \mid Y_k) = \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \exp\left(-\frac{1}{2} \left[\frac{O_s - Y_k}{\sigma_\varepsilon} \right]^2 \right) \tag{1}$$

Where:

- O_s is an observation for one zone at one time
- Y_k is the k^{th} model output
- $L(O_s \mid Y_k)$ is the likelihood of observing O_s given model prediction Y_k
- σ_ε is the measurement error

The total likelihood of a simulation is the sum of the likelihoods of each point prediction:

$$L(O \mid Y_k) = \prod_{s=1}^{S} L(O_s \mid Y_k) \tag{2}$$

where S is the number of measurements.

The Bayes Factor, a ratio of the simulation likelihood to the sum of all the simulation likelihoods, describes how well a simulation fits the data, compared to all of the other simulations. This is computed as:

$$p'_k = \frac{L(O \mid Y_k)\, p(Y_k)}{\Sigma_{i=1}^{U} L(O \mid Y_i)\, p(Y_i)} \tag{3}$$

Where:

- p'_k is the posterior probability of the k^{th} simulation
- $p(Y_k)$ is the prior probability of the k^{th} simulation
- U is the number of Monte Carlo simulations

The posterior probability (Bayes Factor) of each simulation is applied to each parameter associated with that simulation, and thus a re-estimation is obtained for all input uncertainties. This includes both model input parameters, conceptual models, and output predictions. The posterior mean, variance and correlation coefficient are calculated for each input or output parameter and a weighted estimate of the value of the parameter estimated. (Sohn et al 2000)

$$Mean: \quad \mu'_V = \sum_{i=1}^{U} V_i * p'_i \tag{4}$$

$$Variance: \quad \sigma'^2_V = \sum_{i=1}^{U} (V_i - \mu'_V)^2 * p'_i \tag{5}$$

$$Correlation: \quad \frac{\sum_{i=1}^{U} (V_i - \mu'_V)(W_i - \mu'_W) * p'_i}{\sigma'_V * \sigma'_W} \tag{6}$$

Where:

- V and W are any model input or output
- σ'_V and σ'_W are the standard deviations of the model inputs or outputs

Description of House Model

The study house is a five room, one story residential house with an approximate volume of 283 m^3 (Figure 1). Cracks exist in the ceiling of every room, connecting directly to the roof (there is no attic). There is no HVAC system or duct work. The interior Living Room/Kitchen door, the exterior Bathroom window and the exterior Bedroom 1 window are open. The interior Bedroom 2/Living Room door and an exterior Living Room window are randomly varied in the model simulations. Wind blows at a steady 3 m/s perpendicular to one wall of the building.

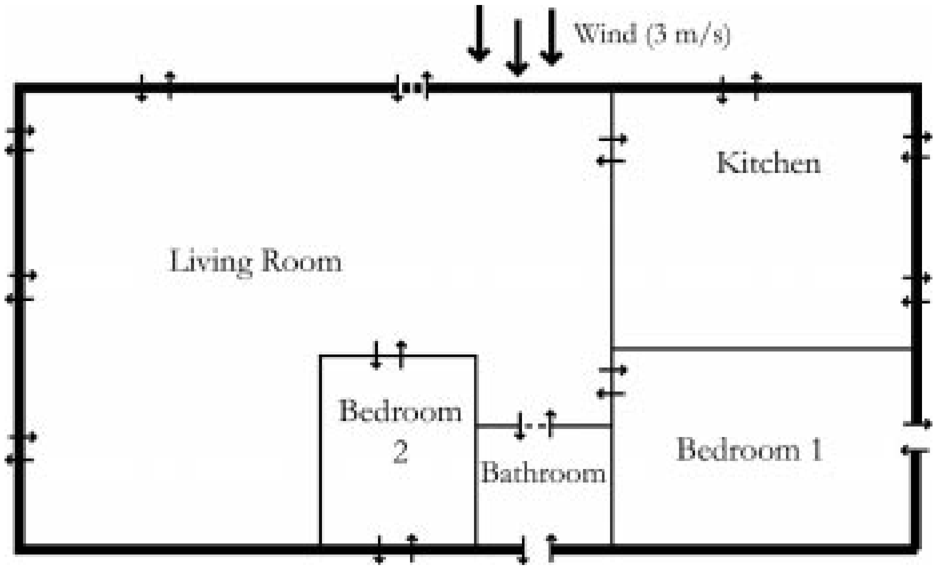

Figure 1: House Floor Plan

Our goal is to locate and characterize the source of a pollutant released in a small building without high quality data describing the source and building. Pollutant transport modeling, however, requires source characterization and building description parameters as inputs to produce pollutant concentration data. We will assume this information to be uncertain and will use the BMC approach to work backwards from pollutant concentration data and low quality building description and source characterization data to reduce uncertainty in input parameters.

Modeling of building airflow and pollutant transport was predicted using the COMIS infiltration model (Feustel et al. 1990). COMIS is a multizone model that assumes each of the rooms behave as well-mixed zones. The COMIS air flow calculation is based on calculating pressure differentials between zones (rooms) through orifices.

Uncertainty Characterization

As discussed earlier, uncertainty distributions were assigned to model input parameters. Table 1 summarizes each of the uncertain input parameters and their assigned uncertainty distribution. Five hundred model predictions were simulated from these distributions. Figure 2 shows an example model prediction for a source released in the living room. Summary statistics of the set of 500 model predictions were calculated. Figure 3 illustrates the median and the two-sided

90% confidence interval from the 500 simulations. The wide interval demonstrates the large uncertainty due to the uncertain source characterization and building description.

Table 1: Input parameters examined in this study and assigned distributions.

Parameter	Distribution	Uncertainty Characterization
Source Location	Random	Rooms 1, 2, 3, 4, 5
Source Mass	Uniform	5 - 100 grams
Source Duration	Uniform	5 - 20 minutes
External Temperature	Uniform	10 - 25°C
Bedroom 2 Interior Door	Random	Open/Closed
One Living Room Window	Random	Open/Closed

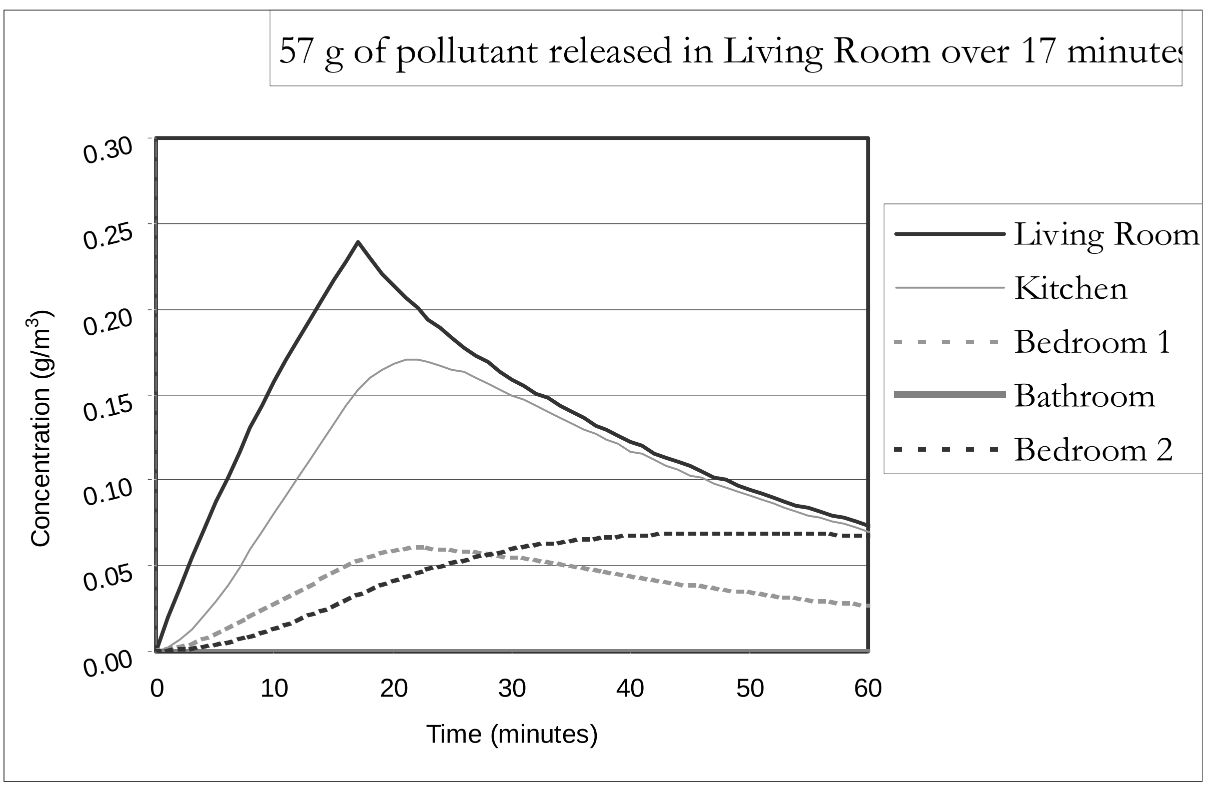

Figure 2: Example of a model prediction of pollutant concentration data.
The concentrations in the bathroom are nearly zero, so are not visible on the figure.

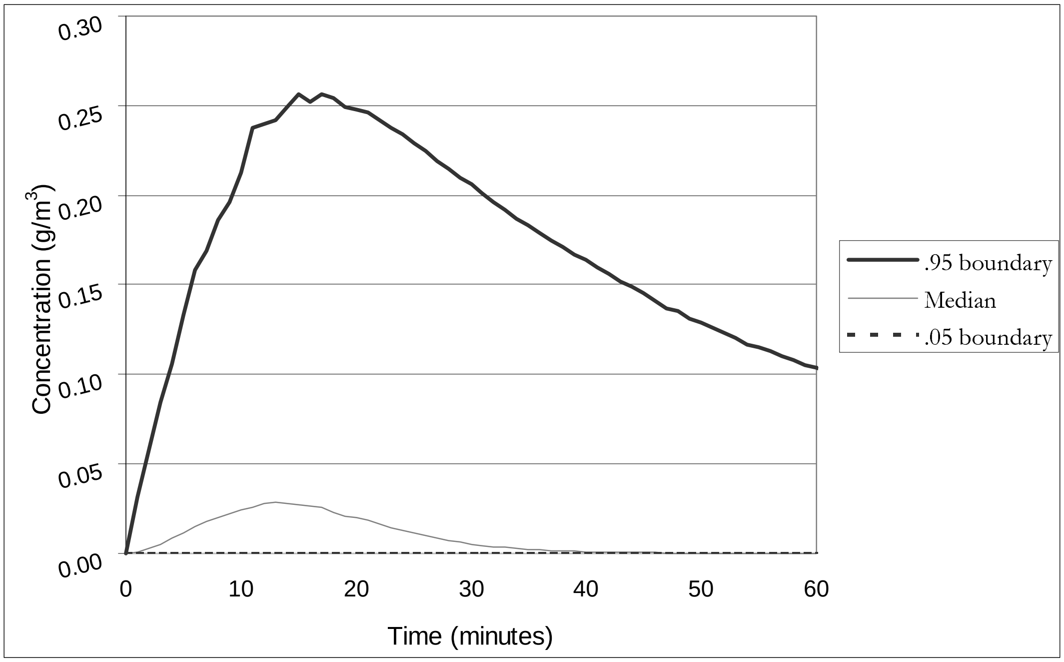

Figure 3: 90% Confidence Bounds for Living Room after 500 simulations

Generation of Synthetic Data

For the Bayes Updating, field data used in this paper was synthetically generated. We chose a simulation using 57 grams of a pollutant released in the Living Room for 17 minutes, with an external temperature of 12.7 °C and both living room window and Bedroom 2 door closed. Hypothetical error was applied to the model simulation by assuming (i.) a normally distributed error associated with a fixed Coefficient of Variation of .10 and (ii.) a normally distributed random error that is independent of the magnitude of the measurement ($\mu = 0$, $\sigma_{random} = .01$ g/m^3). Figure 4 illustrates the synthetic data at five minute intervals.

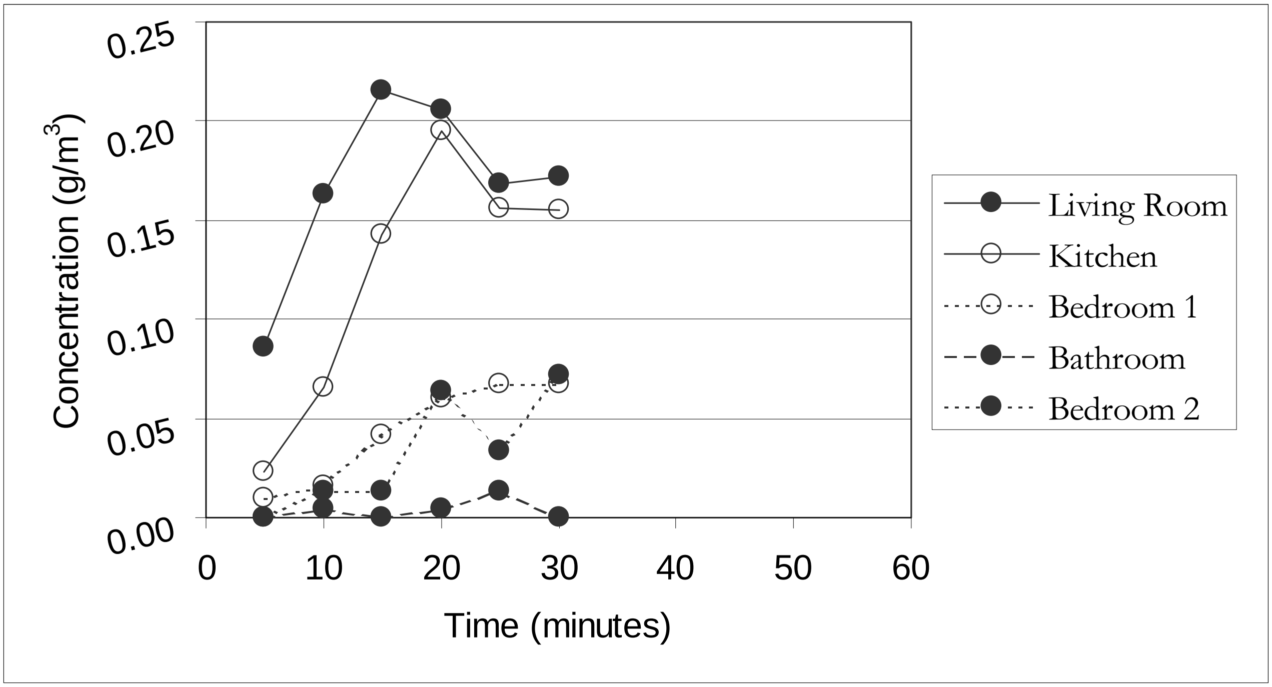

Figure 4: Synthetic Data

Results

Two data gathering scenarios were considered. In the first scenario, we wished to learn if we could locate where the source was released if we had data obtained in each of the five rooms every five minutes. The BMC updating was conducted sequentially as the synthetic data is "obtained". Hence, the first data is obtained at t = 5 minutes and our uncertainty is updated. At t = 10 minutes, when the next sequential data is obtained, new updated uncertainty is estimated. Table 2 summarizes the updated results. Note in Table 2 that uncertainty in input parameters as characterized by standard deviation decreases to at most .1 after 30 measurements are taken. Figure 5 shows the reduction of uncertainty in the source location. Even with just the measurements taken at five minutes, we are able to locate the source in the living room with a probability of 100%. Figure 6, comparing prior to posterior confidence bounds for the Living Room, illustrates the narrowing of the 90% confidence interval at each measurement.

Table 2: Posterior Parameter Uncertainties, after updating using data collected in all rooms, every five minutes. Reduction in uncertainty is reduced, illustrated by decreasing standard deviation (σ). Updated parameter estimations approach correct values.

	Source Mass (grams)		**Source Duration** (minutes)		**External Temperature** (degrees C)		**Living Room Window** probability	**Bedroom 2 Door** probability
	μ	σ	μ	σ	μ	σ		
Prior:	53	27.5	13	4.4	17.5	4.3	open = 50% closed = 50%	open = 50% closed = 50%
Posterior: **5 minutes** (5 measurements total, one in each room.)	47	7.8	14	1.5	13.2	3.0	open: 0 closed = 100%	open: 0 closed = 100%
10 minutes (10 measurements total, one in each room.)	46	4.2	14	0.6	11.8	2.1	open: 0 closed = 100%	open: 0 closed = 100%
15 minutes (15 measurements total, one in each room.)	50	2.3	15	0.5	13.0	2.0	open: 0 closed = 100%	open: 0 closed = 100%
20 minutes (20 measurements total, one in each room.)	51	0.4	15	0.4	12.8	1.1	open: 0 closed = 100%	open: 0 closed = 100%
25 minutes (25 measurements total, one in each room.)	51	0.2	15	0.1	13.3	0.3	open: 0 closed = 100%	open: 0 closed = 100%
30 minutes (30 measurements total, one in each room.)	51	0.1	15	0.0	13.3	0.1	open: 0 closed = 100%	open: 0 closed = 100%
Correct Answer	*57 g*		*17 min*		*12.7°*		*closed*	*closed*

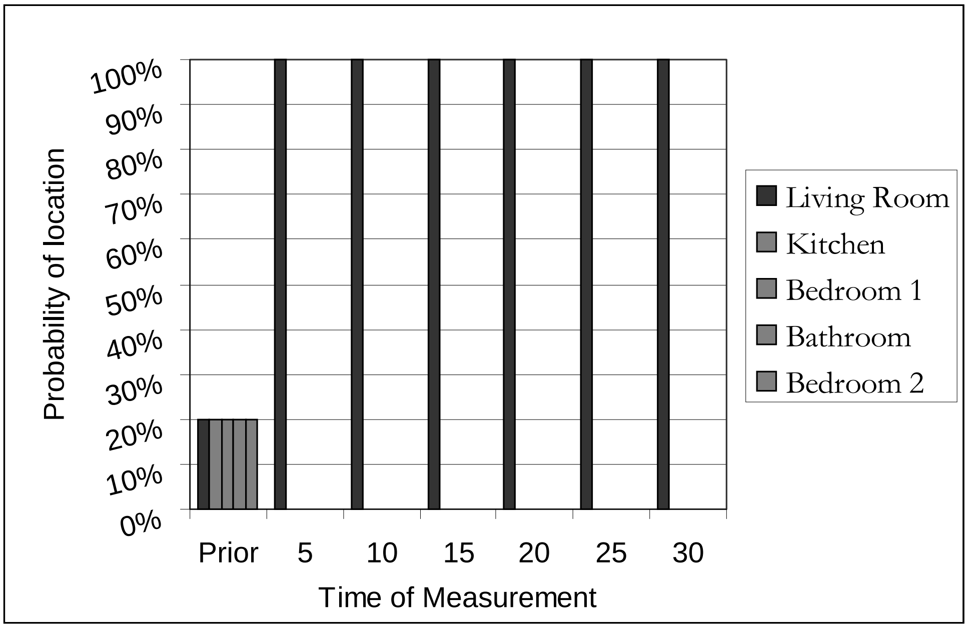

Figure 5: Scenario 1, reduction in uncertainty for source location. Note that after only 5 measurements, at t = 5 minutes, the source has been located in the Living Room.

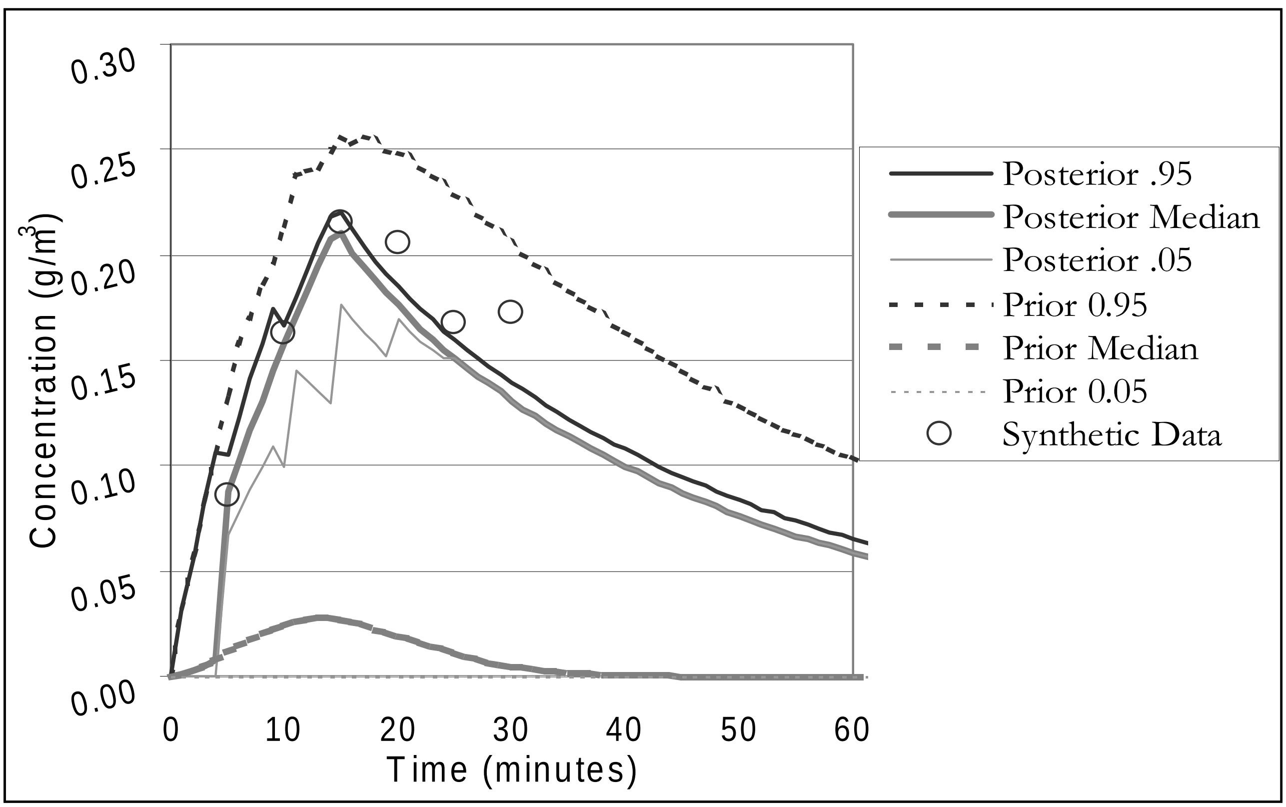

Figure 7: Scenario 1, Prior and Posterior Confidence Bounds for Living Room. The posterior 90% confidence interval narrows as data is received. Concentrations for the Prior .05 bound are near 0.

In the second scenario, we considered data taken in a single room every five minutes. A different room is sampled every five minutes. Uncertainty is significantly reduced at the end of 30 minutes, though only 6 measurements have been used. Yet even after one measurement, the posterior probability of the source being located in the living room is 65%. Table 3 summarizes the results. Figures 7 and 8 illustrate the reduction of uncertainty in the source location and the narrowing of the 90% confidence interval at each measurement, respectively.

Both cases approach the correct values for pollutant mass, duration of release, location, external temperature, and window and door status. In both scenarios the simulation from which the synthetic data was produced is summarized in Tables 2 and 3 as the "correct answer".

Table 3: Posterior Parameter Uncertainties, after updating using data collected in one room every five minutes. Uncertainty is reduced more slowly than Scenario 1. Updated parameter estimations approach correct values.

	Source Mass (grams)		**Source Duration** (min)		**External Temperature** (°C)		**Living Room Window**	**Bedroom 2 Door**
	μ	σ	μ	σ	μ	σ	probability	probability
Prior:	53.0	27.5	13.0	4.4	17.5	4.3	open = 50% closed = 50%	open = 50% closed = 50%
Posterior: **5 minutes** (1 measurement total, latest in Living Room.	65	21.2	12	3.9	17.1	4.4	open = 44% closed = 55%	open = 64% closed = 36%
10 minutes (2 measurements total, latest in Kitchen.	54	19.3	12	4.2	16.9	4.6	open= 44% closed= 56%	open = 52% closed = 48%
15 minutes (3 measurements total, latest in Bedroom 1.	50	16.0	11	3.9	16.4	4.8	open= 35% closed= 65%	open = 51% closed = 49%
20 minutes (4 measurements total, latest in Bathroom.	45	12.0	12	3.8	16.0	4.6	open= 32% closed= 68%	open = 57% closed = 43%
25 minutes (5 measurements total, latest in Bedroom 2.	43	9.2	12	3.0	13.8	4.5	open= 42% closed= 58%	open = 10% closed = 90%
30 minutes (6 measurements total, latest in Living Room	50	2.9	15	0.5	13.6	1.9	open= 0 closed= 100%	open = 0 closed = 100%
Correct Answer	*57 g*		*17 min*		*12.7°*		*closed*	*closed*

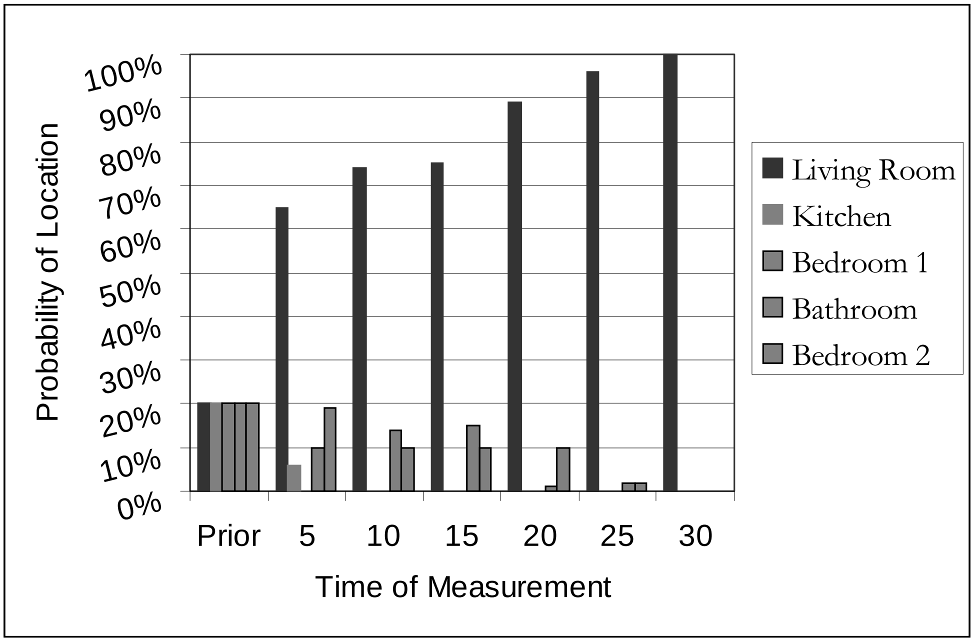

Figure 7: Scenario 2, reduction in uncertainty for source location. Note that reduction of uncertainty in location is slower with data taken in only one room every 5 minutes.

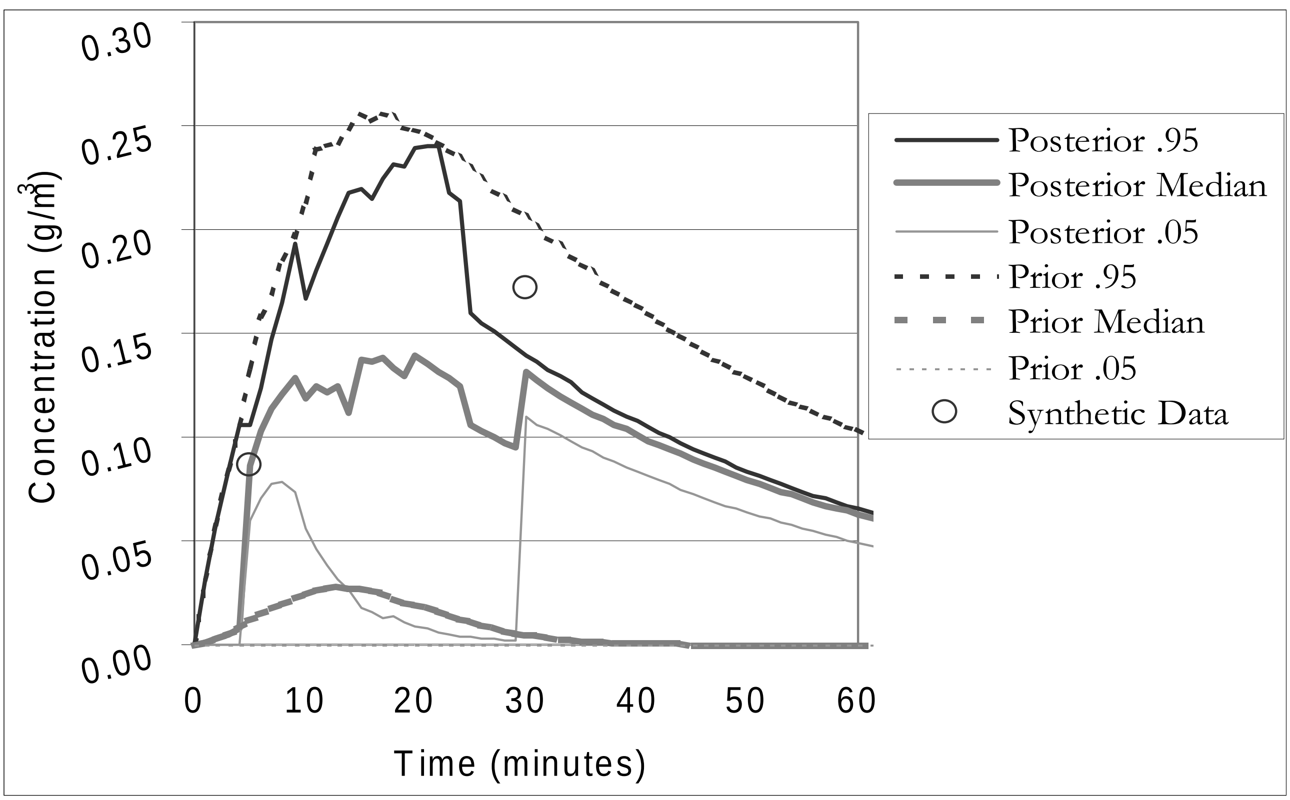

Figure 8: Scenario 2, Prior and Posterior Confidence Bounds for Living Room. The 90% confidence interval narrows more slowly than in Scenario 1. Note that measurements in other rooms are being obtained at 10, 15, 20, 25 minutes. Concentrations for the Prior .05 bound are near 0.

Discussion and Conclusions

After final updating of input parameters, input uncertainty is significantly reduced, even with sparse data (Scenario Two). Reduction in uncertainty is considerable when data is collected in more than one room simultaneously (Scenario One). In both scenarios input parameter distributions are narrowed to the parameter values of the model prediction which most closely matches the simulation from which the synthetic data was produced.

Further research will be required to consider a larger set of uncertain parameters, including additional elements of building description, and to understand how measurement error could effect our updating. Additional research could also include applying the approach using field data from an actual pollutant release in a building.

Acknowledgements

This research was funded in part by the United States Department of Energy under the Energy Research Undergraduate Laboratory Fellowship Program. The research described in this paper was performed at the Ernest Orlando Lawrence Berkeley National Laboratory, a national scientific research laboratory sponsored by the United States Department of Energy.

Many thanks go to my mentor Mike Sohn and to Group Leader Ashok Gadgil. Special thanks go to Jennifer McWilliams for giving me the opportunity to research at LBL, and for her support and assistance. I would also like to thank the members of the Airflow and Pollutant Transport and the Energy Performance of Buildings Groups, my research partner Navtej Singh, and LBL's Center for Science and Engineering Education office. Special thanks to Carlos Figueroa of Cabrillo College for his support.

References

Feustel, H., Allard, F., Dorer, V., Garcia, E., Grosso, M., Herrlin, M., Mingsheng, L., Phaff, H., Utsumi, Y., Yoshino, H. 1990. COMIS Fundamentals. LBL-28560. Lawrence Berkeley National Laboratory.

Sohn, M., Small, M. 1999. "Parameter Estimation of Unknown Air Exchange Rates and Effective Mixing Volumes from Tracer Gas Measurements for Complex Multi-zone Indoor Air Models". Building and Environment. 34: 293-303.

Sohn M., Small M., Pantazidou, M. 2000. "Reducing Uncertainty in Site Characterization using Bayes Monte Carlo Methods". Journal of Environmental Engineering. In press, October 2000.